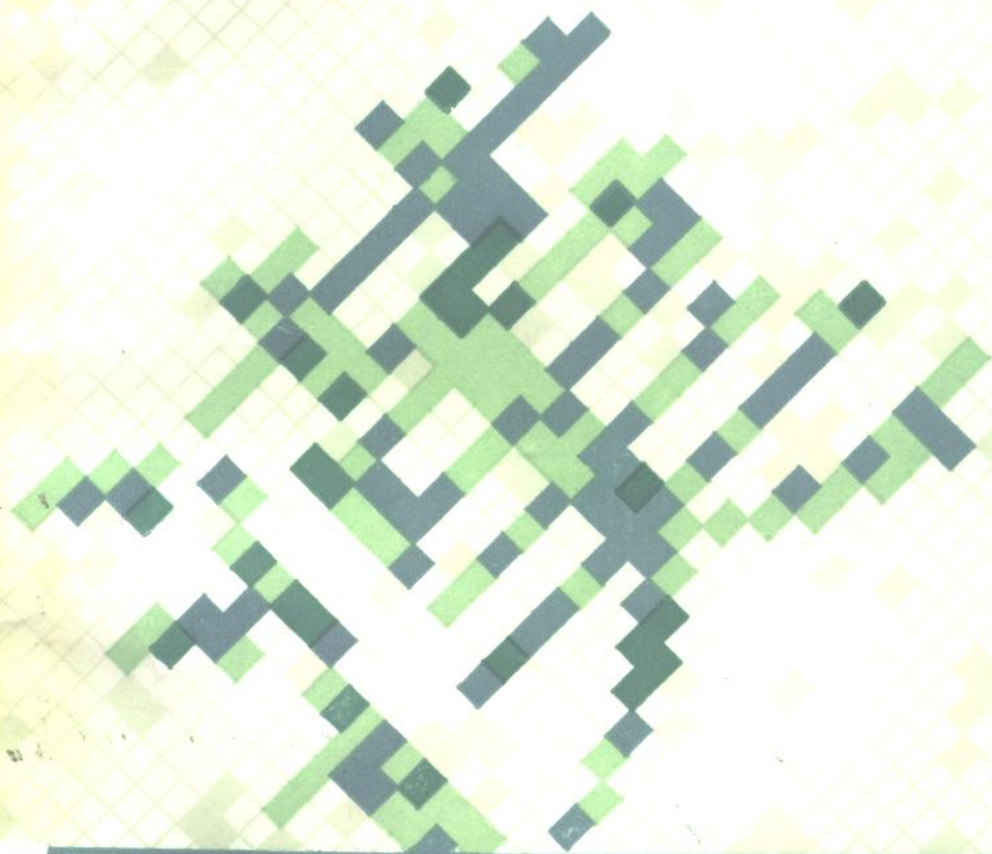


软件的汉化技术

李亦何 编著



CS:AAC0	MOV	CX,0020
CS:AAC3	MUL	CX
CS:AAC5	PUSH	ES
CS:AAC6	PUSH	CS
CS:AAC7	POP	ES
CS:AAC8	ADD	DL,10
CS:AACB	MOV	CS:[AAFF],DL
CS:AAD0	MOV	CS:[AAFD],AX

;每个汉字32个字节
;算出字模首地址

;AX:DX=
;初始化

软件的汉化技术

李亦何 编著

同济大学出版社

内 容 提 要

本书以 IBM-PC/XT、80286/386 等系列的微机环境为背景,系统地叙述了软件的汉化技术、汉化方法和汉字系统的移植要点。全书主要分三个部分,即基础部分、汉化部分和移植部分。本书由浅入深,从汉化工作必须掌握的基本知识和技能开始,分别对汉字系统的功能、运行环境、汉化工作的步骤和方法进行了深入的讨论,对汉字系统的功能扩充、造字、汉字打印和在各种 CRT 显示终端上的移植技术均有新颖而详细的描述。为了使读者更好地理解 and 掌握汉化技术与系统移植原理,书中还给出了许多程序实例供读者学习和参考。

本书可作为高等院校计算机专业高年级学生的参考教材,也可作为从事计算机科学、计算机工程、计算机应用和汉字系统及汉字软件研制、生产、开发的科技人员的参考用书。

责任编辑 郁 峰

封面设计 陈益平

软件的汉化技术

李亦何 编著

同济大学出版社出版

(上海四平路 1239 号)

新华书店上海发行所发行

浙江诸暨印刷厂印刷

开本: 850×1092 1/32 印张: 9.375 字数: 263 千字

1991年7月第1版 1991年7月第1次印刷

印数: 10000 定价: 3.90元

ISBN 7-5608-0774-7/IP·69

前 言

电子计算机的发明，对于人类的科学文化生活，具有划时代的意义。当代社会是充满信息的社会，由于信息量的日趋庞大，信息结构也愈来愈复杂，用计算机实现对信息的处理和管理，已势在必行。但是，电子计算机诞生在西方，其硬件设备，操作系统和大量丰富的软件适用于西文的处理，西文拼音字的字符少，结构简单，处理起来比较方便。而在我国，所使用的文字是汉字，计算机所处理的信息大部分是汉字信息。由于汉字的数量很多，字形结构也较复杂，因此，用计算机处理汉字信息就受到一定程度的限制。为使我国科学技术迅速发展，早日实现现代化，推广计算机的应用是十分必要的。因此，研究和开发汉字信息处理技术具有重大的现实意义。

近年来，我国的计算机工作者作了大量的开发工作，克服了许多技术难关，推出了不少适应处理汉字的硬件设备和汉字操作系统。这些基础工作，为我国计算机应用技术提供了良好的应用环境。因此，在我国的各个领域，计算机的应用范围日益扩大。譬如书刊、情报资料管理系统，文字处理、编辑排版系统，宾馆、医疗卫生管理系统，金融、企业的信息管理和办公自动化等，都离不开计算机管理。除此之外，还有更多的项目迫切地等待着我们去开发和应用，社会需要是技术进步的强大动力，计算机的推广应用不仅是计算机事业发展的目的，也是科学技术水平发展的标志。

在各种计算机应用领域中，均需要有不同的系统软件和应用软件的支持。要在我国使用计算机，只有能够处理汉字信息的应用软件才具有生命力。但是，有许多优秀软件却并不具备处理汉字信息的功能，这就给我们提出了一个问题，就是要把西文软件汉化。或者说，对于用户，要有一个中西文兼容的应用环境。

虽然，我国自行开发了许多好的应用软件，但是，这些还远远不能满足微机应用的社会需求。对于许多国外著名的软件，若不经过汉化处理，就无法在汉字操作系统环境下使用。许多计算机用户往往被动地等待着汉化软件的推出，然后才去予以使用，这样，不仅耽误了软件应用的时间，而且由于一些软件不是自己开发和汉化的，因此，在运行的过程中出了一点小问题，往往就会一筹莫展，这些问题给用户带来了许多不便。因此，汉化技术是目前计算机领域中一个比较重要的课题。可以说，几乎每个从事计算机软件开发和使用的人都会碰到这个问题。从计算机应用的发展眼光来看，不仅系统开发人员应该掌握这一技术，而且应用者也应该了解和掌握汉化技术，这样将会大大减少软件开发的周期和使用故障，并且扩大了用户选择软件环境的余地。由于各种软件之间的结构差异很大，因此，至今没有系统地研究和讨论其规范，也没有出版过这方面的论著。但曾有不少专家提出，应把软件的汉化作为一项软件工程来研究，作者认为是很有必要的。

本书是作者在长期的软件开发工作中积累的资料和多次为开办系统级培训班编写的讲义稿的基础上整理而成的。其旨意在于为读者、尤其是为想在系统作二次开发的初学者提供一些了解系统的思路和软件汉化的技巧。对于高级程序员来说，本书可作为一本编程手册，以便提高系统开发的能力和较快地掌握软件汉化的技术。

书中的第1~3章介绍了汉化技术需要掌握的基础知识，第4、5章描述了汉化工作的具体步骤和关键技术，第6~8章讨论了汉字系统下各种功能的扩充和移植问题，希望能给读者一些有益的启发和帮助，更欢迎计算机同行们共同探讨软件的汉化技术，推动我国计算机事业迅速发展。

在编写本书的过程中，刘杰先生给予了许多热情的帮助，并得到了许多专家的大力支持和鼓励，尤其是在编辑出版时，同济大学出版社的郁峰先生提出了许多诚恳的建议和具体的指导，在此表示衷心感谢。此外，对出版社的有关编辑、审稿、制图人员

的辛勤工作一并表示深切的谢意。

由于汉字系统和汉化技术涉及面广、其对应的软件种类繁多，再加上作者才学浅陋，时间仓促，书中难免会有不少缺点和错误，敬请读者不吝指教。

作者

1990年7月

于上海

目 录

第 1 章 概 论.....	1
1.1 引言.....	1
1.2 汉化软件的概况.....	2
1.3 软件汉化的要求.....	4
1.4 二次开发的过程.....	5
第 2 章 软件汉化的基础	7
2.1 中西文软件的开发环境.....	7
2.2 汉字系统的基本结构.....	8
2.2.1 系统层次及功能.....	8
2.2.2 CC-BIOS 代码系统	10
2.2.3 汉字机内码的设计思想.....	13
2.2.4 代码转换.....	14
2.2.5 汉字编码表.....	17
2.2.6 汉字库与字符点阵.....	19
3 CC-BIOS 基本输入输出功能	20
2.3.1 显示系统的功能.....	20
2.3.2 显示模块的功能调用.....	21
2.3.3 输入系统的功能.....	28
2.3.4 键盘模块的功能调用.....	32
2.3.5 增加新输入编码的方法.....	33
2.3.6 打印系统的功能.....	43
2.3.7 打印模块的功能调用.....	46
2.4 汇编语言指令系统.....	47

2.4.1 寄存器和标志位.....	4
2.4.2 8088/8086 指令系统.....	50
2.4.3 80186/80286 新指令.....	60

第 3 章 汉化软件的工具及其应用..... 63

3.1 工具软件的性能.....	63
3.1.1 动态调试程序.....	63
3.1.2 全屏幕调试程序.....	65
3.1.3 特殊功能调试软件.....	66
3.1.4 反汇编器.....	66
3.2 调试程序的运行状态.....	68
3.2.1 中断方式.....	68
3.2.2 单步跟踪.....	69
3.2.3 断点方式.....	70
3.3 调试程序的应用.....	71
3.3.1 Debug 程序的调用.....	71
3.3.2 Debug 程序使用要点.....	73
3.3.3 Debug 命令.....	74
3.4 程序反汇编的要点.....	87
3.4.1 软件反汇编.....	87
3.4.2 Debug 直接反汇编.....	88
3.4.3 DOS 重定向反汇编.....	89

第 4 章 软件汉化的步骤与方法..... 94

4.1 汉化的基本要点.....	95
4.2 原西文软件的分析方法.....	99
4.2.1 静态分析法.....	99
4.2.2 动态分析法.....	102

4.3 目标程序的跟踪技巧	103
4.4 屏幕显示汉字化	108
4.4.1 屏幕初始化.....	109
4.4.2 中/西文刷屏.....	111
4.4.3 批量数据显示的汉化.....	112
4.4.4 屏幕显示模式的定义.....	114
4.5 全屏幕功能的汉化	115
4.5.1 西文字符显示原理.....	115
4.5.2 调用 CC-BIOS 功能的汉化法.....	118
4.5.3 直接显示汉化法.....	122
4.5.4 下拉式菜单的汉化.....	123
4.6 输入词法逻辑的汉化	125
4.7 字符处理及其定义域的扩展	127
4.7.1 制表符的汉化.....	127
4.7.2 字符域的扩展.....	128
4.8 屏幕显示重新布局	130
4.8.1 光标定位的调整.....	130
4.8.2 屏幕显示信息的调整.....	132
4.9 人机交互信息的汉化	133
4.9.1 查找信息汉化法.....	133
4.9.2 批数据信息汉化法.....	134
4.9.3 数据解释汉化法.....	136

第 5 章 软件汉化的实施 138

5.1 dBASE-III Plus 的汉化	138
5.1.1 显示输出的汉化.....	138
5.1.2 变量名及输入部分的汉化.....	145
5.1.3 “滤除符”和提示信息的汉化.....	148
5.2 Turbo-Pascal 的汉化	149

5.3 Lotus 1-2-3 显示模块的汉化	153
5.4 增添程序模块的方法	156
5.4.1 命令文件的扩充	157
5.4.2 执行文件的扩充	158
5.4.3 用中断方法扩充程序	161
第 6 章 汉字库及造字原理	164
6.1 扩展内存读写汉字库的技术	164
6.1.1 保护方式寻址原理及功能	164
6.1.2 全局描述器表 (GDT)	166
6.1.3 保护方式汉字模块读写的实现	169
6.2 汉字字形及其结构	174
6.2.1 汉字库的存储	174
6.2.2 汉字字形结构	176
6.2.3 高点阵字形结构	177
6.3 造字系统的功能	179
6.3.1 高点阵的汉字显示	179
6.3.2 造字的原理和设计	182
6.3.3 造字网格的显示	186
6.3.4 改码及磁盘读写	188
6.4 汉卡系统造字的考虑	190
第 7 章 汉字打印驱动程序的设计	194
7.1 打印机汉字打印原理	194
7.2 汉字打印变倍方法	196
7.2.1 变换运算无级变倍法	197
7.2.2 变换字形的整形和补偿	199
7.2.3 笔划平滑放大法	201
7.2.4 表检索无级变倍法	203

7.3 汉字打印驱动程序	208
7.3.1 环境设备的兼容性	209
7.3.2 数据结构和编程	212
7.4 汉字打印系统死锁及其解决方法	218
第 8 章 汉字系统的环境及其移植	224
8.1 显示器概述	224
8.2 CGA 显示卡	226
8.2.1 CGA 卡的基本结构	226
8.2.2 CGA 编程要点	229
8.2.3 汉字显示改进	232
8.3 MDA 单色显示卡	236
8.3.1 MDA 卡的基本结构	236
8.3.2 MDA 显示卡编程要点	238
8.3.3 MDA 汉字显示模块的移植	241
8.3.4 MDA 仿真 CGA 显示	243
8.4 CGE400 显示卡	245
8.4.1 Color400 卡的基本结构	245
8.4.2 Color400 卡的编程要点	250
8.4.3 汉字显示模块移植要点	255
8.5 EGA/VGA 显示卡	258
8.5.1 结构与控制原理	258
8.5.2 EGA 寄存器设置	261
8.5.3 EGA 显示器编程要点	266
8.5.4 EGA/VGA 汉字显示模块的移植	272
附录 I 二、十、十六进制数转换表	277
附录 II 功能扩展键扫描码表	280
参考文献	284

第1章 概 论

1.1 引言

目前，电子计算机的应用已经深入到各个领域。我国是使用汉字的国家，计算机技术无论应用在哪个领域，几乎全离不开汉字信息的处理。因此，为使计算机技术在我国能得到更广泛的应用和更迅速的发展，就必须开发大量能够处理汉字的系统软件和应用软件。如何使计算机软件能够处理中文信息呢？从语言文字学的角度来看，需要一种专门的中文信息处理系统。若从世界科技文化的共性来看，则需要通用的信息处理系统。由此说来，我们应该采用“中西文兼容”的方法，在世界范围内交流信息。怎样实现这种“兼容”技术？“兼容”到何种程度？这是我国广大计算机专业人员非常关注的问题。本书就是针对这一问题展开讨论。

多年来，许多计算机专业人员致力于研究计算机的汉字处理方法，根据计算机的硬件设备以及系统软件和应用软件的使用条件和环境进行了大量的二次开发工作，使我国计算机的应用得以广泛地普及和推广。譬如：CC-DOS 汉字操作系统、桌面字处理、轻印刷排版系统和适应于各个层次的汉字输入编码等等，都说明了计算机在处理汉字信息方面取得了许多成果。然而，系统软件仅仅给广大用户建立了一个汉字系统的环境。在实际应用中，绝大多数用户需要各种类型的应用软件包，如各种高级语言或数据库等软件，来开发和设计适应于本系统或本部门的生产、科研、教学和事务处理等方面的应用系统。尤其像 IBM-PC/XT、80286/80386 等系列的微机，拥有丰富的应用软件包，世界最著名的几家软件公司，为这一系列的微机开发出许许多多的软件。

但是，这么多的软件中，几乎没有一套软件能够比较全面地处理含有汉字的数据信息。它们必须经过不同程度的“汉化”技术处理后，才能处理汉字信息。因此，在我国，汉化软件已经是计算机应用、推广和普及过程中的一种社会需求，也是现阶段我国计算机应用领域中的一个重要课题。

1.2 汉化软件的概况

从我国目前的微机应用状况来看，的确需要快速地组织计算机专业高级人员来开发出大批量的、能够处理汉字信息的实用软件产品并投入市场，以适应或满足汉字软件的社会需求。例如，我国自行开发的 OFFICE, AutodBASE 等软件包，虽然有一定的市场，但从开发周期、数量和使用范围的角度看，还是远远不能满足我国现阶段计算机应用的进程。因为开发一套系统级的软件，需要花费大量的时间、人力和财力。这种方法似乎可行，但却不符合实际。

当然，也有一条捷径：将一些优秀和著名的软件引进后进行汉化，也就是说，对其进行二次开发。这样，可使软件开发的工作量大大减少，既实用、经济，又行之有效。这种做法加快了软件产品推出的周期，能够较好地满足我国计算机迅速发展的社会需求。

在丰富的应用软件集中，有许多是世界公认的优秀产品，例如：Ashton-Tate 公司的 dBASE、Borland 公司的 Turbo 系列和 Microsoft 软件公司的各种高级语言、商口软件和字处理软件等等。这些著名的软件不仅具有一定的代表性，而且，从人类思维的共性来看，其中绝大部分功能是我们能够接受的。这些软件凝聚了许多专家的高技术思想结晶和优良的程序设计结构，直接影响着世界计算机应用水平的发展。虽然处理西文信息与中文信息在某种程度上存在着差异，但就这些软件的最终功能来说，它们是完全能够融合于我国计算机应用系统之中的。

由于计算机软硬件不断地发展和迅速地更新换代。汉化技术至今仍没有一定的标准模式。开发者大都凭借自己的经验；在各种类型的系统环境下进行不同程度的实施。例如屏幕显示卡，从早期 IBM-PC 机的 CGA 的中分辨率，发展到 MDA、Color 400 显示卡和高分辨多功能 EGA/VGA 显示卡。软件方面也从行编辑、全屏幕编辑发展到各种下拉式菜单的多窗口软件。在软硬件迅速更新换代的过程中，许多专业人员作了大量的工作，结合具体环境，对各种系统软件和应用软件进行了汉化工作，以满足各行各业计算机应用的需要。这对于推动我国计算机应用水平的提高和发展起着很大的促进作用。但是，纵观我国整个计算机界软件的汉化情况，尚有许多不足之处，主要有如下几个方面：

1. 缺乏软件汉化的理论指导；
2. 汉化技术未普及，用户较被动；
3. 从软件工程角度来说，缺乏对汉化软件的统一评价标准；
4. 许多汉化后的软件整体结构有缺陷；
5. 没有真正地达到中西文兼容。

上述不足之处受多种因素的限制和影响。因为，无论是系统软件还是应用软件，它们的种类是多种多样的，所使用的环境亦各有差异，各种软件中有关阻碍汉字信息处理的模块不同，其逻辑地址也不同。另一方面，由于所要汉化的软件大都是在目标代码下进行分析和改造的，汉化者基本上是根据经验和本身的技术素质来作二次系统开发工作，各人对计算机系统和指令系统掌握的深浅以及对原西文软件分析的粗细，均直接影响汉化后软件产品的质量。

汉化一个软件，从原理上讲，并不十分复杂，就是把原西文软件或某些代码文件作适当的调整和部分修改，使之能够处理汉字信息。但是具体实施起来，则并非易事。这要求专业人员掌握、了解许多计算机系统软硬件的基础知识、汉字操作系统的功能和工具软件的使用等等。对于初学者来说，一时很难下手。他们应该在对计算机的内部结构、外设接口和汉字操作系统等了解的同

时，多读一些源程序，收集有关的资料，这对于下一步的汉化工作将会有很大裨益。本书对于汉化技术的基础知识也作了一些介绍，并由浅入深地讨论了汉化软件的环境、过程及其方法。

1.3 软件汉化的要求

软件的汉化与软件的分析不同，前者主要是寻找软件中与汉字信息处理有关的部分，对其进行流程分析，针对不同的问题作出不同的修改方案。后者是全方位系统地分析软件的设计思想、结构及其内部逻辑关系。

对于汉化软件这项软件工程来说，我们的最终目的应该是：西文软件能够运行的环境，汉化后的软件同样可以运行，汉字系统的工作环境应该是包含着西文软件同样的工作环境。汉化后的应用软件应不受输入输出系统和操作系统版本的限制，换句话说，就是有较强的兼容性和通用性，在此基础上，达到汉化软件的最高境地。通常，对于一个汉化后的软件的评测标准和要求有如下几项：

1. 对原西文软件的代码变动应尽可能的少，在不丢失原西文软件的功能条件下，取得最佳的汉化效果；
2. 对西文软件内的有关模块实行功能强化，使其兼顾到汉字处理的特点；
3. 对西文软件内的有关子程序实行功能转化，使其从单一处理西文字母转化为对中西文均能处理；
4. 人机界面的信息汉字化。

前三个项目作为一个汉化级别，意味着汉化后的软件已经具备了识别汉字信息的功能，扩展了程序对于字符处理的定义域。第四项作为二级汉化，是针对中西文系统环境下软件所具备的不同的信息吞吐量间的匹配而进行的，并且使人机之间的交流更接近日常语言。

任何一个软件，无论是版本的差异或是功能的差异，其汉化

的原理和目标都是相同的，最终的结果也是一致的。但是，通过分析研究，我们可以看到，各种软件需要修改的程序模块以及与汉字信息有关的部分，它们在内存中运行时的逻辑地址均不相同，现阶段还没有一个非常有效的智能型的汉化软件的工具，绝大多数都是采用 Debug 或 PCTOOLS 等一些动态调试程序针对具体的软件进行汉化工作，经过反复的修改调试，最终成为可以处理汉字信息的软件产品，其方式如图 1-1 所示。

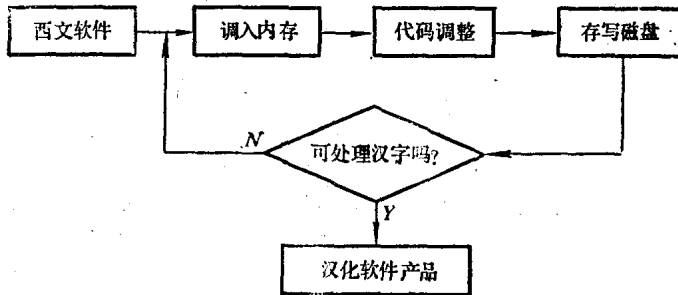


图 1-1 软件的汉化流程

西文软件大多数是在显示器字符方式下运行，尤其是一些字处理软件，所显示的字符由显示卡的字符发生器产生并显示。而汉字系统，如 CC-DOS，是在图形方式下处理汉字信息的。仅从 CRT 屏幕显示这一例子，就说明它无法运行。即使有些汉字系统的环境能够接受字符方式正常显示，但是，其定义的汉字内码也将与原西文软件中的 ASCII 码的定义发生冲突。因此，如果一个西文软件要在汉字系统环境下正常处理汉字信息的话，就必须进行汉化处理，这是唯一的前提。

1.4 二次开发的过程

设计一个软件，是将解决问题的思维过程定义翻译为一种软件表示的过程，而汉化一个软件，则是对该软件表示过程的扩充。因此，在标准软件的规范下完成评测软件汉化的优良程度，

也需要一些原则，才能顺利通过二次开发的进程，这些原则如下所列：

1. 分析原软件的层次组织，以便弄清软件元素之间的控制关系；
 2. 在原软件的部分逻辑上，实现处理中文信息的子功能；
 3. 从总体上逐级求精，用不同的手段进行各个层次的测试，最终完善为软件产品；
 4. 建立二次开发的文档和资料，便于今后的系统维护。
- 图 1-2 所示说明了二次开发的过程。

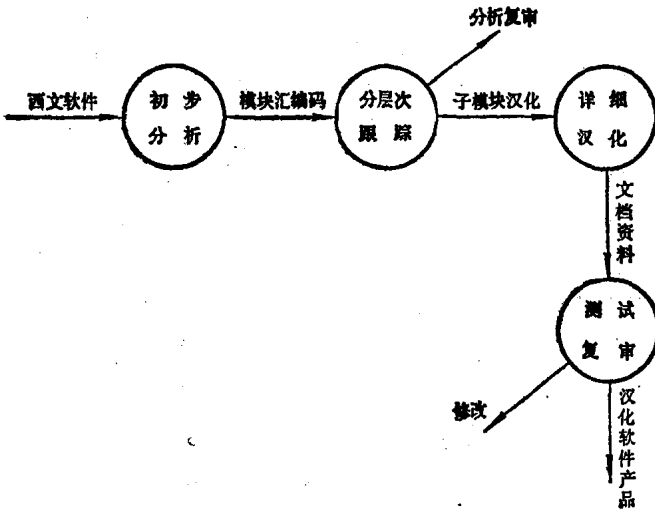


图 1-2 二次开发的过程

在当今世界计算机技术迅速发展的形势下，无论是系统程序员还是软件应用者，都应该尽快熟练地掌握软件的汉化技术。这将有助于系统的升级和软件版本的不断更新，有力地推动我国计算机应用事业的迅速发展。